D10.2 – Research Data Management Plan, version 3.0

Realising an Applied Gaming Eco-system

Research and Innovation Action

Grant agreement no.: 644187

D10.2 – Research Data Management Plan, version 3.0

RAGE – WP10 – D10.2

Project Number	H2020-ICT-2014-1
Due Date	Month 48, January 2019
Actual Date	Month 48, January 2019
Document Author/s	Eric Kluijfhout (editor), Christina Steiner, Michael Kickmeier-Rust, Sabina Guaylupo, Andrea Molinari, Rubén Riestra, Matthias Hemmje
Version	3.0
Dissemination level	PU/ RE
Status	Final
Document approved by	Wim Westera

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644187

Document Version Control			
Version	**Date**	**Change Made (and if appropriate reason for change)**	**Commentator(s) or Author(s)**
1.1	18-12-2016	Version distributed to the co-authors for first review	Eric Kluijfhout
1.2	18-01-2017	Feedback processed	Sabina Guaylupo, Wim Westera, Boyan Bontchev, Heike Görzig
1.3	26-01-2017	Research metadata approach added. DMP-version for internal use to guide data collection for the round-1 pilots	Michael Kickmeier-Rust
2.0	26-01-2018	Checked for required updates; minor text adaptations	Eric Kluijfhout
3.0	23-01-2019	Version 3 updates on research data sets and research metadata approach, minor text adaptations throughout the document	Christina Steiner

Document Change Commentator or Author		
Initials	**Name**	**Institution**
E.	Kluijfhout	OUNL
C.	Steiner	TU Graz
M.	Kickmeier-Rust	TU Graz
S.	Guaylupo	Inmark
A.	Molinari	Okkam
R.	Riestra	Inmark
M.	Hemmje	FTK
B.	Bontchev	Sofia University
H.	Görzig	FTK
W.	Westera	OUNL

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY

What data will be collected, processed or generated during the RAGE project? Following what methodology and standards? And what data will be shared and/or made openly available, and how will it be curated and preserved? These issues are typically described in a Data Management Plan (DMP), outlining how research data is handled during the project and after the project is completed.

The RAGE DMP specifically provides guidelines on ethics, data protection and open research data access to RAGE researchers involved in WP5 (Case experiments) and WP8 (Validation). Ethics and data protection are especially relevant in view of the games and audiences targeted by RAGE. Therefore, RAGE is one of the participating projects in the EU open research data pilot, an initiative under Horizon 2020 that aims at improving and maximising the access to and re-use of research data created in European projects.

This document is the third and final iteration of the DMP. The first version was published in month 6 of the project, an intermediate version was created mid of the project, and the version 3 update was done towards the end of the project. The DMP thus is not a fixed document; it has evolved and gained more precision and substance during the lifespan of the project.

This final version of the DMP includes updates that are informed by the following project results:

- The *D8.3 - First RAGE Evaluation Report* (delivered in November 2017) presents the results of the formative evaluations on the RAGE components and Ecosystem portal and the validation studies in the context of the application scenarios (pilot round 1 – see also *D5.1 - Scenario Arrangement document – round 1* and *D5.3 – Pilots quality report round 1)*. It details on the evaluation instruments actually applied and data sets collected in the different evaluation tasks.
- The milestone document *MS16 – Second pilot validation instruments* (delivered in March 2018) presents an update on the procedures and tools to be used in the second round of pilot evaluations.
- The *D8.4 – Second RAGE Evaluation Report* (delivered in December 2018) (delivered in December 2018) presents the results of the summative evaluations on the RAGE components and Ecosystem portal and the validation studies in the context of the application scenarios (pilot round 2 – see also *D5.2 - Scenario Arrangement document – round 2* and *D5.3 – Pilots quality report round 2)*. It details on the evaluation instruments actually applied and data sets collected in the different evaluation tasks.

1 INTRODUCTION

RAGE participates in the EU open research data pilot and as such will deliver a Research Data Management Plan (D10.2). Open access to research data refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user.

The DMP is not a fixed document; it evolved and gained more precision and substance during the lifespan of the project.

Version 1 of the DMP was delivered in month 6 and outlined the preliminary data management approach and foreseen datasets collected or generated by the project. An intermediate version 1.3 was created in January 2017 which further specified and updated the data sets to be collected as part of the first round of pilot evaluations (see section 2.1, Applied games data sets) planned to start early 2017, and addressed a number of aspects of the research data management approach that were left open in version 1 (see section 2.2.1). The final version is prepared towards the end of the project, incorporating updates resulting from the outcomes of the work in the context of the evaluation and validation studies. This document version provides updates of the data sets generated as part of the pilot evaluations (see section 2.1) and of the research data metadata formats (see section 3).

1.1 The EU open research data pilot

References to research data management are included in Article 29.3 of the Model Grant Agreement:

> *Regarding the digital research data generated in the action ('data'), the beneficiaries must:*
>
> *(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
>
> *(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
>
> *(ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan*

According to the EU open research data pilot, a Data Management Plan[1] (DMP) describes the data management life cycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

Each dataset that will be generated by the project has to be described in terms of five dimensions in compliance with the template provided by the Commission (see Annex 1):

1. Data set reference and name: a unique persistent identifier for the data set
2. Data set description: a description of the data set which specifies the origin, scope, scale, beneficiaries and the link to the corresponding publications (if any).
3. Standards and metadata: a reference to relevant standards and a description of the metadata schema adopted to describe the data.
4. Data sharing: all the information concerning access and reuse of the dataset including the nature of access (open or restricted), the tools or software needed, the reference and type of the repository where data are stored.
5. Archiving and preservation (including storage and backup): long-term preservation procedures, costs and volume of preserved data.

[1] From: Guidelines on Data Management in Horizon 2020, Version 1.0, 11 December 2013

The DMP describes the data management life cycle of research data. Open access to the publications based on these research data (the 'golden road' to open access through Open Access publications, or the 'green road' through access to repositories and/or self-archiving[2]) are out of scope for the DMP and are described in the RAGE Dissemination Plan.

1.2 The RAGE data management approach

In addition to the above EU-requirements on data sets, the RAGE Grant Agreement[3] outlines the project's data management approach, comprising fifteen elements. These fifteen elements not only address the data management lifecycle but also address how to meet recent EU-guidelines and European and national legislation on ethics and data protection.

The three organizing principles are:

- **Ethics**, where the guiding principles are informed consent, privacy, voluntary participation, and (personal) data ownership (the right to change or delete personal data).
- **Data protection**, where the guiding principles are security in collection, storage, retrieval and destruction of data; anonymized personal data and confidentiality.
- **Open access**, where the guiding principle is that research data should be/remain available to fellow researchers free of charge for validation and (re)use.

The scope of this final version of the DMP builds on a number of other project deliverables:

- D8.3 *'First RAGE Evaluation Report',* which was delivered in month 33 by Technische Universität Graz. This deliverable describes presents the results of the formative evaluations on the RAGE components and Ecosystem portal and the validation studies in the context of the application scenarios (pilot round 1). It details on the evaluation instruments actually applied and data sets collected in the different evaluation tasks.
- D5.1 '*Scenario Arrangement document - round 1*' and D5.3 *'Pilots quality report round* 1' were delivered in month 21 and month 32 by Okkam. These documents outline the implementation plans and, respectively, the actual implementation details for the first round of RAGE pilots, including details related to design, methodology instruments, data policy and privacy policy quality assurance in the different application scenarios.
- MS16 *'Second Pilot Validation Instruments'*, providing an update on the procedures and tools to be used in the second round of pilot evaluations. The predecessor document (MS8) also included a chapter on the use of Zenodo.
- D8.4 *'Second RAGE Evaluation Report',* which was delivered in month 47 by Technische Universität Graz. This deliverable describes presents the results of the summative evaluations on the RAGE components and Ecosystem portal and the validation studies in the context of the application scenarios (pilot round 2). It details on the evaluation instruments actually applied and data sets collected in the different evaluation tasks.
- D5.2 '*Scenario Arrangement document - round 2*' and D5.3 *'Pilots quality report round* 2' were delivered in month 37 and month 47 by Okkam. These documents outline the implementation plans and, respectively, the actual implementation details for the first round of RAGE pilots, including details related to design, methodology instruments, data policy and privacy policy quality assurance in the different application scenarios.

[2] From: http://www.openaccess.nl/whatisopenaccess
[3] Annex 1 - Description of Action Part B, page 128

2 THE RAGE DATA MANAGEMENT PLAN

This final version of the DMP (month 48) is structured in the same way as the first version, comprising of two sections:

- The RAGE data management template (section 2.1). The template lists the research data sets in terms of their five dimensions as outlined in Annex 1. These data sets relate to the three main evaluation objects (Assets, Applied games, and Ecosystem). This section has been updated in accordance with the data sets generated and collected as part of the pilot evaluations, as compared to the data sets initially specified in this DMP.
- Implementing the RAGE research data management approach (section 2.2). This lists the fifteen principles of the project's data management approach and how and when these will be implemented. The principles are addressed by the adoption of OpenAIRE-Zenodo as the preferred open access repository to manage the project's research data and the guidelines of D8.1 '*RAGE Evaluation Framework and Guidelines'*.

2.1 RAGE data management templates

Evaluation and validation of the RAGE results focuses on three groups of data sets, as follows:[4]:

- **Assets (game components):**
  - Usability
  - Software quality
- **Applied games:**
  - Educational effectiveness:
    - Usability
    - Game experience
    - Learning effectiveness
    - Transfer effect
  - Pedagogical costs and benefits
- **Ecosystem services and processes:**
  - Quality
  - Benefit

D8.1 '*RAGE Evaluation Framework and Guidelines*' further elaborates the evaluation questions, evaluation criteria, and methods for collecting and analyzing data to holistically evaluate the RAGE technologies.

Data set: Assets – Usability	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on the ease with which functional software assets can be used (understood, learned, used and are considered attractive) by a game developer in creating a game, and data on their

[4] Annex 1 - Description of Action, p. 35

	acceptability (relevance) - whether and how the software assets will be used in the real game development world[5]. Data are collected through (online) questionnaires, structured observations and prototyping. Data collection is based on both shared instruments and independent local instruments. Scientific publications based on these data and their authors[6] will be referenced here as well.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8' community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

Data set: Assets - Software quality	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on the quality of the functional software assets, with a focus on the added value and effect on the game development process. This includes the perceived usefulness for game development, the impact of asset integration on game engineering, asset functionality, perceived benefits and cost effectiveness (costs and benefits) for game development and applied games. Data are collected through questionnaires, structured observations, prototyping, or interviews/focus groups. Data collection is based on both shared instruments and independent local instruments.

[5]Bevan, N., Kirakowski, J., Maissel., J. What is Usability? Proc. of the 4th Int. Conf. on HCI, Stuttgart, Germany, Sept. 1991

[6] E.g. through unique global IDs like ORCID IDs

	Scientific publications based on these data and their authors will be referenced here as well.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8' community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

Data set: Applied games – Educational effectiveness	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on the educational effectiveness of applied games for learners and training providers. This covers the following aspects: ○ Usability: Are users able to interact easily with the applied games? ○ Game experience: How do end users experience the use of the applied games? (e.g. in terms of enjoyment, flow, usefulness for learning) ○ Learning effectiveness: Do the applied games effectively support learning? ○ Transfer effect: Do the applied games support transfer of acquired knowledge/skills to the performance context? Data is collected through a mix of evaluation instruments, combining qualitative and quantitative data (e.g. open answers and ratings) as well as subjective and objective measures (e.g. self-reports and actual test performance). Importantly, also a combination of post- or retrospective assessments (e.g. questionnaire feedback after game session) with continuous data or in-game data (e.g. observations, game-based user

	data) is used. The mixed-method approach in evaluation instruments provides the possibility of data triangulation and a more comprehensive understanding of educational effectiveness. Scientific publications based on these data and their authors[7] will be referenced here as well.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8' community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

Data set: Applied games – Pedagogical costs and benefits	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on experienced or expected benefits and additional costs of the application and integration of RAGE applied games in training. The main evaluation and data collection instruments consist in surveys and semi-structured interviews with training developers. Data collection is based on shared instruments across pilot studies and independent local instruments.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8'

[7] E.g. through unique global IDs like ORCID IDs

	community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

Data set: Ecosystem services and processes – quality	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on the quality of the ecosystem services and processes for different stakeholder groups in the context of applied games, in terms of e.g. usability, system performance and features, import/export quality, and tutorial quality. Data are collected through questionnaires, ratings, prototyping, or interviews/focus groups. Data collection is primarily based on shared instruments. In addition, from common logging and tracking data gathered and managed by the ecosystem, relevant data on the usage of the ecosystem will be explored and used for evaluation purposes. Scientific publications based on these data and their authors will be referenced here as well.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8' community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and

	the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

Data set: Ecosystem services and processes – benefit	
Reference and name (identifier)	Once the data set is created and stored in the Zenodo open access repository, its generated Digital Object Identifier assigned by Zenodo will be published.
Description	Evaluation data collected on users' perception of ecosystem services and processes in terms of perceived added value, usefulness and user acceptance. Data will be collected through explicit data collection via questionnaires, prototyping, or interviews/focus groups with shared instruments. In addition, usage data logged by the ecosystem will be exploited for evaluation purposes and complement subjective evaluation feedback. Scientific publications based on these data and their authors will be referenced here as well.
Standards and metadata	To enable fellow researchers to search and access the open data sets produced in RAGE, the consideration of metadata standards as well as (quasi) standardized vocabularies is crucial. RAGE will provide a stakeholder-centered framework for metadata descriptions of RAGE data sets on the basis of existing standards such as Dublin Core and established classification approaches such as those of the ACM and the APA. This vocabulary will be implemented for the 'RAGE WP8' community in the OpenAire-Zenodo open access repository by TUGraz. A detailed description of the approach is provided in chapter 3.
Data sharing	The research and evaluation data is managed in the OpenAire-Zenodo open access repository by TUGraz, for which it has created and manages the 'RAGE WP8' community in Zenodo. The data access procedures and technical mechanisms for dissemination are defined by Zenodo's functionality. Zenodo exposes its data to OpenAIRE, helping researchers to comply with the Open Access demands from the EC and the ERCs. The data controller will award the appropriate (creative commons or closed) license to each data set, depending on the sensitivity of the data (preferably anonymized data with an open access license).
Archiving and preservation	The research data will be managed in the Zenodo open access repository, for which the 'RAGE WP8' community has been created which is curated by the partner leading the evaluation work package (TU Graz, WP8).

2.2 Implementing the RAGE research data management approach

2.2.1 RAGE research data management approach and its implementation

The RAGE Grant Agreement outlines fifteen elements as part of its data management approach. Below we list each of them together with their proposed implementation strategy.

#	RAGE DM principle	Description[8]	Proposed implementation actions	Implementation
1	RAGE EDP-group	Within RAGE we will install an internal ethics and data protection group for identifying mechanisms for handling personal data properly and for the alignment across different pilots.	Form the RAGE ethics and data protection group comprising Eric Kluijfhout, Rubén Riestra, Christina Steiner (after version 1 replaced by Michael Kickmeier-Rust), Sabina Guaylupo, Andrea Molinari and Matthias Hemmje. This group is responsible for compiling the first version of the DMP delivered in month 6.	Realised in M6
2	Localised responsibility	We will opt for a decentralised approach, since the particular approaches and preferences to ethics and data protection is at the discretion of the separate research partners. Every out of 6 local pilots will be assigned a local responsible researcher to take into account local ethical guidelines and data protection procedures as well as transcending EU guidelines and regulations.	Each research institute involved in a pilot and/or empirical evaluation assigns a national responsible researcher. As a first task, this researcher will collect information on institutional and national ethical guidelines and data protection procedures and share this. Next, the responsible researchers will discuss their findings. The preliminary outcomes will serve as input to the *RAGE Evaluation Framework and Guidelines* (D8.1, month 12).	Covered in D8.1, annex A2.5
3	Localised procedures	The pilots will conform to local/institutional procedures and approvals by local/institutional ethics and data protection authorities. Signed copies of approvals will be timely made available to the European Commission.	The local responsible researcher submits the pilot research design to the respective institutional and/or national authorities, and upon approval make this available to the European Commission. Ethical approvals will be collected and stored by the data protection group. The pilots mainly relate to T8.5 *Validation studies in application scenarios'*, but this principle relates to all WP8 tasks in which data is collected. Defining the exact procedure within the context of the pilot design and execution is a shared responsibility between WP5 (Task 5.1: *Aligning the approaches to application scenario arrangement*) and WP8 (Task 8.1:	Covered in D8.1, annex A2, D5.1-D5.4

[8] See Annex 1 - Description of the Action Part B, page 128

			RAGE Evaluation Framework and Guidelines).	
4	Data collection instruments	The collection of research data will be based on both shared instruments and independent local instruments.	Guidelines on data collection instruments will be formulated as part of D8.1 *RAGE Evaluation Framework and Guidelines* (month 12). Data collection instruments are selected in line with the evaluation questions addressed and the evaluation setting. Where possible, shared instruments have preference. Where available and suitable, standardized and/or established instruments are used.	Covered and reported in D8.1, annexes 3/4/5; MS8, MS16, D8.3 and D8.4
5	Data protection at various levels	RAGE aims to develop good practice and will comply with national and institutional data protection initiatives that aim to harmonise the approaches and systems (e.g. DANS).	The local responsible researchers' will investigate the present status of their local institutional and national research data repositories (e.g. do they exist; are they recognized as a trusted digital repository by e.g. TrustedDigitalRepository.eu; can they be located through a registry service like re3data.org). They will share this information to come up with a proposed RAGE policy on institutional/national/international data storage and protection procedures and tools for decision making by the SMB.	Covered in D5.1-D5.4, Procedures and guidelines are covered in D8.1 and MS8, including a brief manual on how to use Zenodo
6	Participant personal data ownership	All data are owned by the participant, who is entitled to withdraw his/her data from the sample at any time.	The option to withdraw data from the research data set will be included in the Information Letter. Contact details of the data management service will be included in all information channels to the pilot participants. Mechanisms for retrieving and deleting participant data will be devised and described in D8.1.	Information letter and consent form templates included in D8.1, annex A2.3

7	Complaints service	RAGE will arrange a complaints service for participants.	This will be a multi-staged service: located at the institutional (the research institute) level, with the option to escalate to the project level (Technische Universität Graz as responsible for the evaluation and validation activities). At the institutional level it will tie in with existing institutional and national privacy procedures. Contact details will be included in all information channels to the pilot participants.	Covered in D8.1, annex 2: the localized structure will be used (see 2 and 3 above), and communicated through the Information letter
8	Information letter	In the Information Letter RAGE will clearly communicate its research intentions, its ambition to make the anonymised research data openly available, and its eventual commercial perspectives.	Based on the information letter consent from evaluation participants outside the project will be collected (in case of minors also from their parents). From project members participating in evaluation protection/confidentiality of data collected from consortium partners is guided by the non-disclosure guidelines in the Consortium Agreement without the need for extra consent collection.	Information letter and consent form templates included in D8.1, annex A2.3
9	Open data sharing	RAGE anticipates to use relational database CSV-like formats for open data sharing.	This will apply to data sets deposited in local/institutional repositories as well as national and international ones.	Zenodo allows the submission of all sorts of data formats, and selection of appropriate licenses to be awarded to data sets. Depending on the sensitivity this may be a creative commons or closed license.
10	Technical data protection	Participant data will be subjected to technical data protection procedures (collection, storage, retrieval and destruction) in accordance with national and EU legislation.	The ‘national responsible researchers’ will include the issue of data protection procedures in their analysis of the local/national status (as part of #5). Local research data repositories need to be identifiable through a register of research data repositories like Re3data, and be certified by the TrustedDigitalRepository.eu.	Covered by the selection of Zenodo.

11	Data storage location	Common tracking and logging data needed for running the games are stored and managed by the game companies in the secure user databases of the games' delivery platforms. All user data originating from research data collection instruments (e.g. online questionnaires) are securely stored and managed during the research by the responsible local research institute.	The game companies will provide information on the security of their user databases (location of the hosting services, backup and recovery procedures) to the 'national responsible researchers' who will assess these against national guidelines and legislation. RAGE will draw up requirements/protocol for storing and managing user data at the local research institute as part of D8.1.	Covered in D8.1, section 3.3 'Working with the Results – Processing, Storing, and Sharing Data'
12	Anonymised data	Anonymity of collected and pre-existing user data is effected as a first step after data collection. Personal user data are replaced with a unique identifier, while coded tables remain available for restoring the original dataset.	A unified approach will be drawn up, outlining the main procedural steps and management of the identifier list.	Covered in D8.1, section 3.3.3.2 'How to Anonymize'
13	Ecosystem users	User data collected from the Ecosystem will be stored and managed securely and remain under the full control of the end-users.	This is a requirement for the selection and configuration of the ecosystem-tools in WP6. The data will also be used for evaluation purposes.	>M24
14	Research data preservation	Most scientific journals require the preservation of research data for 5 years or longer after the publication of a paper. Therefore, destruction of research data will not take place in the near future, unless the participant would request for this.	Needs no further elaboration.	Covered by the selection of Zenodo.
15	Audio- and video recordings	Since it is hard to anonymise audio and video recordings, individuals will be explicitly informed about such recordings, their purpose, privacy protection and usage conditions/intentions. In principle research recordings will be destructed after processing unless otherwise agreed. Also audio or video recordings for publicity purposes will always go	Needs no further elaboration.	Information letter included in D8.1, annex A2.3

		with detailed prior information and informed consent.		

With respect to the DM Principles nº2 Localised responsibility, nº 3 localised procedures and nº 5 data protection a various levels, the RAGE consortium and especially the Data Management Group is aware that the current legal status in the Member States is changing.

We are aware that on April 2017 the EC published the official text of the GDPR, a new single law is working on a new single law on Data Protection. This new regulation, among other points, aims to update and modernise the principles enshrined in the 1995 Data Protection Directive and doing away with the current fragmentation among Member States on this issue.

The new rules become applicable within two years after its publication, i.e. 25 May 2018. The RAGE consortium will take the appropriate measures in order to comply with the new situation. Meanwhile and regarding especially to DM Principle nº 6 Participant data ownership, nº 8 Information letter, and nº 10 Technical data protection, the current Directive 95/46/EC applies as well as national legislation.

2.2.2 RAGE data management use case

The fifteen data management principles are elaborated in this use case:

The educational provider (WP5) defines the target audience and the educational requirements for the pilot game. The game company (WP4) develops the game, while the educational provider delivers its educational content. Ethics and data protection are explicit – non-educational – requirements in the game design and its underlying data management procedures and tools.

The educational provider recruits pilot participants from the target audience, respecting ethic requirements (informed consent, voluntary participation etc.). The research institute (WP8) provides detailed information to the educational providers on purpose and implications of the pilots, which these providers use in their recruitment of participants; sets up the research structure around the validation studies by defining and creating data collection tools and procedures, manuals, etc.; and trains staff of the educational provider on research data collection during pilot execution. Ethics and data protection are explicit requirements in the research design and its procedures and tools.

In both the game design and the research design the frame of reference for ethics and data management comprises national/institutional and transcending European guidelines and legislation. An assigned local (national) researcher is responsible for their correct implementation in the pilot, and their approval by the local/institutional ethics and data protection authorities. A signed copy of the approval will be made available to the European Commission.

The educational provider conducts the pilot. User data are collected a) as part of the game design/scenario, and b) additionally and specifically for research purposes. The final research data set may combine both (in-game data and additionally collected research data). Anonymity of collected and pre-existing user data is effected as a first step after data collection.

The research data set is securely stored and managed by the local research institute. This is guided by ethics-, data protection-, and open access requirements from the local research institute and national and EU-legislation and guidelines on harmonization of data management approaches and systems like OpenAIRE-Zenodo.

3 RAGE RESEARCH DATA METADATA FORMATS

The aim of participating in the open data initiative is to allow fellow researchers to benefit from the recorded data sets of RAGE, to realise further impact of these data sets. Thus, when describing the research data we need to acknowledge the terminology of potential stakeholder groups. Consequently, referring to standards and common metadata description schemes is inevitable.

Metadata is structured information about an object (data) that facilitates functions associated with the object. (Greenberg, 2009[9]). Jane Greenberg (2012[10]) compared a number of metadata approaches for describing research data. She summarized three levels of descriptive schemes and their characteristics:

- Simple: Interoperable, easy to generate/low barrier, generally multidisciplinary, genera/format agnostics, primarily flat, general (not granular), 15-25 properties

- Simple/moderate: Interoperability balanced w/specific needs, generation requires more expertise, greater domain focus, extensible--via connecting to other schemes, more granular, more properties

- Complex: Interoperable level, generation requires expertise, genera focus/format variation, hierarchical, granular, and extensive (100+ properties)

In RAGE, we will take an approach by relying both on simple schemes (preferably Dublin Core Metadata Element Set, http://www.dublincore.org/documents/dces), and complex schemes (preferably the Data Documentation Initiative - DDI, http://www.ddialliance.org). While the metadata scheme provides the scaffolding for data description, we need to integrate such a standard with well-defined vocabularies of the stakeholder groups:

- The *computer science and (game) development* perspective, for which we integrate the classification system and vocabulary of the ACM (https://cran.r-project.org/web/classifications/ACM.html). This offers a hierarchical, well-elaborated classification of topics.

- The *psychological/social* science perspective, for which we build upon the American Psychological Association (APA; http://www.apa.org/pubs/databases/training/class-codes.aspx). RAGE's own *games-specific vocabulary*, the RAGE Applied Gaming Classification System (RAGCS). This was developed for use in the ecosystem, tailored to RAGE's scientific- and practitioners communities. For the purpose of research data metadata, in particular the categories 'target groups' and 'skills' from RAGCS are relevant – to characterise the subjects/evaluation participants involved in the data collection and the knowledge domain or type of skills targeted.

The *learning and competence development* perspective, for which it is not trivial to find a common understanding of related terms. As one example, the term competency (which is vital in this context) has a sheer infinite number of meanings, ranging from a rather atomic element of any aptitude to broad and complex determinant of task performance. In this regard, in addition to the skill categorisation in line with RAGCS we will take ESCO (pillar 'skills/competences'[11]) as a reference framework to describe the learning and competence dimension of the research data related to the pilot studies in the application scenarios. ESCO is a multilingual classification system covering skills, competences, qualifications and occupations and provides a common

[9] Greenberg, J., White, H.C., Carrier, S., & Scherle, R. (2009). A metadata best practice for a scientific data repository. Journal of Library Metadata 9 (3-4), 194-212.
[10] Greenberg, J. (2012). Metadata for Managing Scientific Research Data. Webinar of SILS Metadata Research Center: http://www.niso.org/news/events/2012/dcmi/scientific_data/
[11] European Union (2017). ESCO handbook. European Skills, Competences, Qualifications and Occupations. https://ec.europa.eu/esco/portal/document/en/0a89839c-098d-4e34-846c-54cbd5684d24

reference terminology to describe the targeted learning objectives and outcomes of the applied games.

These foundations are used to build a controlled vocabulary for describing the RAGE research data in Zenodo, comprising a term list, hierarchies, and a thesaurus (cf. Hedden, 2010 for this process[12]). A practical approach is taken in order to enable manageable data curation and to capture all relevant information for data (re)use, while avoiding providing unnecessary information. Conceptually the description scheme is structured as given below. This structure is used for the description of the research data complementing the default metadata fields foreseen in Zenodo for data upload:

- Topic description (the 'aboutness' of data)
  - ACM CCS 2012 classification[13]
  - PsycINFO classification categories and codes[14]
- Name entities
  - Organizational information
  - Geographic information
  - Time information
- Data Description
  - Types (from plain text to standardized media formats)
  - RAGCS target group
  - Evaluation dimensions
    - Evaluation object
    - Methodology/design
    - Evaluation variables
  - Instruments
  - Knowledge/skill elements
    - RAGCS skills
    - ESCO skills classification
  - Relationships (to related resources, such as other data sets, publications)

To easily generate the metadata for the data sets to be published in Zenodo, a template covering the relevant metadata fields and providing guidance is used as a basis for publishing the research data in the Zenodo open access repository.

[12] Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. Journal of Digital Asset Management (2010) 6, 279-284.
[13] https://dl.acm.org/ccs/ccs_flat.cfm
[14] https://www.apa.org/pubs/databases/training/class-codes.aspx

ANNEX 1: DMP TEMPLATE FOR EU OPEN RESEARCH DATA

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but evolves during the lifespan of the project.

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

- ***Data set reference and name***

  Identifier for the data set to be produced.

- ***Data set description***

  Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

- ***Standards and metadata***

  Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

- ***Data sharing***

  Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

  In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, and security-related).

- ***Archiving and preservation (including storage and backup)***

  Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.